PRince: a web server for structural and physicochemical analysis of Protein-RNA interface

Amita Barik, Abhishek Mishra and Ranjit Prasad Bahadur*

Department of Biotechnology, Indian Institute of Technology, Kharagpur 721302, India

ABSTRACT

We have developed a web server, PRince, which analyzes the structural features and physicochemical properties of the protein–RNA interface. Users need to submit a PDB file containing the atomic coordinates of both the protein and the RNA molecules in complex form (in '.pdb' format). They should also mention the chain identifiers of interacting protein and RNA molecules. The size of the protein–RNA interface is estimated by measuring the solvent accessible surface area buried in contact. For a given protein–RNA complex, PRince calculates structural, physicochemical and hydration properties of the interacting surfaces. All these parameters generated by the server are presented in a tabular format. The interacting surfaces can also be visualized with software plug-in like Jmol. In addition, the output files containing the list of the atomic coordinates of the interacting protein, RNA and interface water molecules can be downloaded. The parameters generated by PRince are novel, and users can correlate them with the experimentally determined biophysical and biochemical parameters for better understanding the specificity of the protein–RNA recognition process. This server will be continuously upgraded to include more parameters. PRince is publicly accessible and free for use. Available at http://www.facweb.iitkgp.ernet.in/~rbahadur/prince/home.html.

INTRODUCTION

Protein–RNA interaction is ubiquitous in many cellular processes. The interaction is specific in nature, and non-specific interaction can lead to malfunction of the cell. Interfaces formed by interactions between protein and RNA molecules provide context for understanding the principles of molecular recognition *in vivo*. Over the last few decades, remarkable progress has been made in understanding the structural and the functional aspect of the protein–RNA interactions from their three-dimensional atomic structures (1–8). The ever expanding Protein Data Bank (PDB) (9), which is the central repository of structural information of the macromolecules and their complexes, also helps in such understanding. Concurrently, there have been attempts to analyze the structural geometry and physicochemical properties of the interfaces using a number of parameters based on the features of the interacting surfaces (10–20). These analyses have been further extended to develop softwares or web servers or both for automatic calculations of these parameters (21–23). Some of these softwares and web servers are used in the prediction of RNA binding sites in proteins (24–29). Nevertheless, in spite of all these developments, our understanding of the protein–RNA recognition process is still not adequate enough to explain the structural basis of the conformational changes during the recognition processes (30), mechanism of sequence specific recognition (31–33), as well as the prediction of protein–RNA complexes through the docking methods (34–36).

In recent years, we have developed several parameters based on the geometric structure and physicochemical properties of the interacting surfaces in biomolecules (37). We have also studied the hydration pattern of the interfaces, and developed parameters to investigate the role of interface water molecules in the recognition process (38). All these parameters are useful in understanding the structural specificity of the recognition process in protein–protein complexes, and are extensively used in discriminating the specific protein–protein interfaces from the non-specific ones (39–42). They have also been successfully used to understand the specific recognition of the RNA molecules on the protein surface, suggesting that the protein–RNA recognition process involves elements of shape recognition as well as electrostatic interaction and the recognition of the base sequence (5,19). In order to calculate all these interface parameters for a given protein–RNA complex, we have automatized the method and implemented it into a web server named PRince (Protein-RNA interface; http://www.facweb.iitkgp.ernet.

*To whom correspondence should be addressed. Tel: +91 3222 283790; Fax: +91 3222 27870; Email: r.bahadur@hijli.iitkgp.ernet.in; ranjitp_bahadur@yahoo.com

in/~rbahadur/prince/home.html). This article describes the development of the web server PRince.

PROGRAM DESCRIPTION

Input file and format

This server allows the users to submit a protein–RNA complex file or a dataset of protein–RNA complexes in the PDB format containing the protein and the RNA chains. Users must also indicate the chain identifiers for each of the protein and the RNA unit (a maximum of eight protein chains and eight RNA chains are allowed). The server can handle up to 20 000 atoms for each of the protein and the RNA chain. The detailed information about the user submitted PDB files and chain identifiers for the protein and the RNA will be displayed on the server page once the calculations are completed.

### Output files and parameters

The server generates four types of output files in PDB format: (i) list of the interface amino acids with their atomic coordinates (1DFU_P.int); (ii) list of the interface nucleotides with their atomic coordinates (1DFU_R.int); (iii) list of the interface atoms on the surface of the interacting protein–RNA complex (1DFU_int.pdb); (iv) list of the interface water molecules along with the interface protein and RNA atoms (1DFU_wat.int). In all these files, 1DFU, the PDB identifier for a given protein–RNA complex, has been used as an example. These files do not contain occupancy factor and B-factor columns; instead they have three columns, in which the solvent accessible surface areas (SASAs) of the constituent atoms in the individual subunit, in the complex and their difference, are provided. Users can download these output files for further calculations; for example, to calculate the interface area contributed by individual residues, or to find out the important water molecules at the binding surface. In addition, they can also display these files by Jmol plugin. Figure 1 shows interface atoms, surface atoms and interface water molecules for the protein and the RNA subunits involved in a complex formed by ribosomal protein L25 and 5S rRNA fragment (PDB id 1DFU). Beside these downloadable files, the server also generates a downloadable table with the statistics of the interface parameters (1DFU_param.txt), which are discussed below.

Statistics of the interface parameters

The size of the protein-RNA interface is estimated by calculating the interface area (B). It is calculated in terms of the SASA of the protein and the RNA molecules and is given by the following equation

$$B = SASA_{protein} + SASA_{RNA} - SASA_{complex} \qquad (1)$$

$SASA_{protein}$ and $SASA_{RNA}$ are the SASA of two interacting molecules and $SASA_{complex}$ is the same of the complex. The interface area B is the area of the protein and the RNA solvent accessible surfaces that becomes buried when two molecules associate. The SASA values are
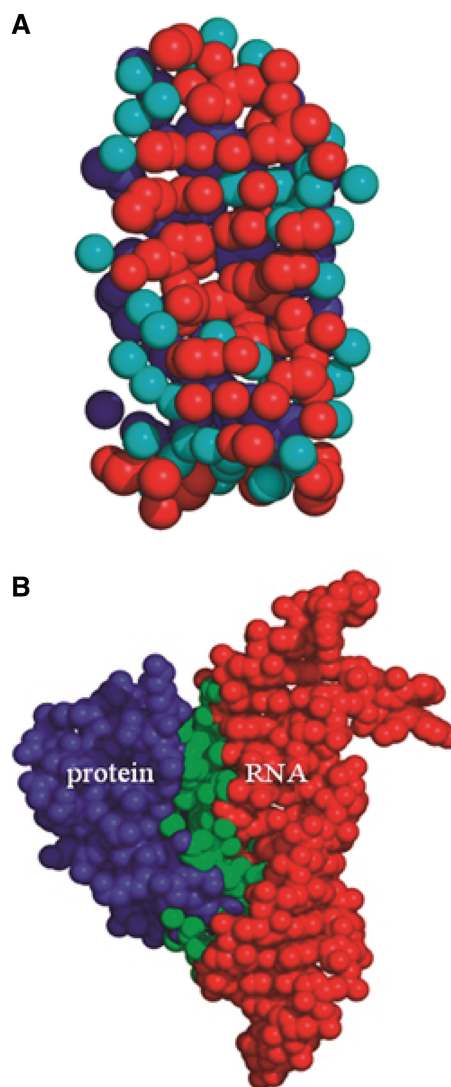
Figure 1. Graphical representation of the protein-RNA interface formed by ribosomal protein L25 and 5S rRNA fragment (PDB id, 1DFU). (**A**) Interface region along with the water molecules. Atoms belong to protein, RNA and interface water molecules are colored blue, red and cyan, respectively. (**B**) Surface region of the protein and RNA are colored in blue and red, respectively, while their interface region is colored in green.

calculated from the atomic coordinates by rolling a solvent probe (with the radius of a water molecule) over the surfaces of the protein and the RNA molecules using the program NACCESS (43), which implements the Lee and Richards (44) algorithm. All the atoms (belongs to amino acid residues and nucleotides) that loose solvent accessibility in the complex and contribute to B are considered as interface atoms. The ratio of the interface area to the rest of the surface area is calculated for the individual protein and RNA molecules as well as for the whole complex.

The chemical composition of the interface or the solvent accessible surface is estimated by measuring the contribution of the different atom types to the respective interface

area and solvent accessible surface area (SASA), and is calculated by the following equation:

$$f_{comp} = \text{interface (or solvent accessible surface)}$$
$$\text{area contributed by a particular atom type/total}$$
$$\text{interface (or solvent accessible surface) area}$$
(2)

The composition is divided into four different types: (i) nonpolar, (ii) neutral polar, (iii) negatively charged and (iv) positively charged. At the protein surface, all the carbon-containing groups are considered as nonpolar; O, N and S are considered as neutral polar; N is positively charged in Arg/Lys side chains; O is negatively charged in Asp/Glu side chains. At the RNA surface, all the carbon-containing groups are considered as non-polar; N and O are neutral polar except O1P and O2P, which are considered as negatively charged (19).

Depending on their spatial distribution, the interface atoms can be divided into two different categories. Those which are not accessible by any solvent molecules are called fully buried interface atoms, and those which are partly accessible to solvent molecules are called partially buried interface atoms (39). The atomic packing of the interface is quantified by the following equation

$$f_{bu} = \text{number of fully buried interface atoms/total}$$
$$\text{number of interface atoms}$$
(3)

This fraction will be higher for a closely packed interface compared to a loosely packed interface.

Local atomic density (LD) index is used to measure the overall density of the interface as described by Bahadur *et al.* (39). In brief, for each interface atom i, the number n_i of the interface atoms that are within a distance 12 Å of atom i in the same subunit is counted. LD is the average of n_i over all N interface atoms and is given by the following equation

$$LD = \sum_i (n_i)/N$$
(4)

LD measures the packing density at each point of the interface.

Polar interactions made by the amino acids and the nucleotides are expressed by the intermolecular hydrogen bonds (H-bonds) between protein and RNA. Water molecules at the interface play an important role in stabilizing the protein-RNA interface by making polar interactions. A water molecule is selected as interface water if it is within 4.5 Å distance from at least one atom of both protein and RNA chains (38). Direct protein–RNA intermolecular H-bonds as well as water mediated H-bonds across the interface are calculated by the program HBPLUS with default parameters (45). All these interface parameters generated by the server for a protein–RNA complex formed by ribosomal protein L25 and 5S rRNA fragment (PDB id, 1DFU) are given in Table 1.

PROGRAM IMPLEMENTATION

The server runs on a 3.0 GHz Xeon processor with Linux operating system. The programs for calculating the

Table 1. Interface parameters for a protein–RNA complex formed by ribosomal protein L25 and 5S rRNA fragment (PDB id, 1DFU)

Parameters	Protein	RNA	Complex
Interface area B (Å^2)	828.2	859.7	1687.9
Surface area buried (%)	14.0	12.0	14.8
Number of interface			
Atoms	84	88	172
Amino acids	26	–	26
Nucleotides	–	19	19
B (Å^2) per			
Amino acid	31.9	–	31.9
Nucleotide	–	45.2	45.2
Fraction of buried atoms	0.30	0.32	0.31
Local density	41	39	40
Interface composition[a]			
Nonpolar	47.4	21.3	34.1
Neutral polar	23.2	45.1	34.4
Negatively charged	4.7	33.7	19.4
Positively charged	24.7	–	12.1
Surface composition[a]			
Nonpolar	55.3	28.7	40.5
Neutral polar	25.4	43.6	35.6
Negatively charged	9.7	27.7	19.7
Positively charged	9.5	–	4.2
Water molecules			
Number per interface	–	–	49
Number per 1000 Å^2	–	–	29
Bridging water[b]	–	–	10
Hydrogen bonds			
Protein–RNA			
Number per interface	–	–	17
Number per 1000 Å^2	–	–	10.1
Water–Protein			
Number per interface	–	–	21
Number per 1000 Å^2	–	–	12.40
Water–RNA			
Number per interface	–	–	71
Number per 1000 Å^2	–	–	42.1

[a]Percentage compositions are calculated using Equation (2) described in the text.
[b]The bridging waters are identified as those interface waters making H-bond with both protein and RNA atoms.

parameters are written in 'C' programming language and the web interface has been developed using Javascript and PHP. The server generally takes 30 s to print the output data for an average size protein–RNA complex, and it can take around a minute for a large complex. Users must install a java plugin or Java Runtime Environment (JRE) for the browser to view the 3D structures of the interfaces in Jmol. This web server is best viewed in Mozilla Firefox, Google Chrome, Safari or Opera, and it may run slow on the older versions of Internet Explorer. In addition, we have provided a non-redundant dataset of 81 protein–RNA complexes and their interfaces compiled by Bahadur *et al.* (19), which can be downloaded from this server website.

CONCLUSION

We have developed a web server PRince, to analyze the structural and physicochemical properties of the protein-RNA interfaces. Users can submit a protein–RNA

complex file with a list of interacting protein and RNA chains. The server generates several parameters describing the structural specificity of the interaction. These parameters could be used for further analysis, and users can correlate them with the experimentally determined biophysical and biochemical parameters for better understanding the specificity of the protein–RNA recognition process. This server will be continuously upgraded to include more parameters. PRince is free and open to all and there is no login requirement for the users.

REFERENCES

1. Cusack,S. (1999) RNA-protein complexes. *Curr. Opin. Struct. Biol.*, **9**, 66–73.
2. Draper,D.E. (1999) Themes in RNA-protein recognition. *J. Mol. Biol.*, **293**, 255–270.
3. Jones,S., Daley,D., Luscombe,N., Berman,H.M. and Thornton,J.M. (2001) Protein–RNA interactions: a structural analysis. *Nucleic Acids Res.*, **29**, 943–954.
4. Chen,Y. and Varani,G. (2005) Protein families and RNA recognition. *FEBS J.*, **272**, 2088–2097.
5. Janin,J. and Bahadur,R.P. (2008) The structural basis of protein-nucleic acid recognition. *Cell. Mol. Bioeng.*, **1**, 327–338.
6. Puton,T., Kozlowski,L., Tuszynska,I., Rother,K. and Bujnicki,J.M. (2011) Computational methods for prediction of protein-RNA interactions. *J. Struct. Biol.*, October 12 (doi:10.1016/j.jsb.2011.10.001; epub ahead of print).
7. Zhao,H., Yang,Y. and Zhou,Y. (2011) Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Res.*, **39**, 3017–3025.
8. Iwakiri,J., Tateishi,H., Chakraborty,A., Patil,P. and Kenmochi,N. (2012) Dissecting the protein-RNA interface: the role of protein surface shapes and RNA secondary structures in protein-RNA recognition. *Nucleic Acids Res.*, **40**, 3299–3306.
9. Berman,H.M., Battistuz,T., Bhat,T.N., Bluhm,W.F., Bourne,P.E., Burkhardt,K., Feng,Z., Gilliland,G.L., Iype,L., Jain,S. *et al.* (2002) The Protein Data Bank. *Acta Crystallogr Sect.*, **D58**, 899–907.
10. Lustig,B., Arora,S. and Jernigan,R.L. (1997) RNA base–amino acid interaction strengths derived from structures and sequences. *Nucleic Acids Res.*, **25**, 2562–2565.
11. Nadassy,K., Wodak,S. and Janin,J. (1999) Structural features of protein–nucleic acid recognition sites. *Biochemistry*, **38**, 1999–2017.
12. Allers,J. and Shamoo,Y. (2001) Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.*, **311**, 75–86.
13. Treger,M. and Westhof,E. (2001) Statistical analysis of atomic contacts at RNA-protein interfaces. *J. Mol. Recogn.*, **14**, 199–214.
14. Chen,Y., Kortemme,T., Robertson,T., Baker,D. and Varani,G. (2004) A new hydrogen-bonding potential for the design of protein–RNA interactions predicts specific contacts and discriminates decoys. *Nucleic Acids Res.*, **32**, 5147–5162.
15. Morozova,N., Allers,J., Myers,J. and Shamoo,Y. (2006) Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics*, **22**, 2746–2752.
16. Baker,C.M. and Grant,G.H. (2007) Role of aromatic amino acids in protein-nucleic acid recognition. *Biopolymers*, **85**, 456–470.
17. Ellis,J.J., Broom,M. and Jones,S. (2007) Protein–RNA interactions: structural analysis and functional classes. *Proteins*, **66**, 903–911.
18. Zheng,S., Robertson,T.A. and Varani,G. (2007) A knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. *FEBS J.*, **274**, 6378–6391.
19. Bahadur,R.P., Zacharias,M. and Janin,J. (2008) Dissecting protein–RNA recognition sites. *Nucleic Acids Res.*, **36**, 2705–2716.
20. Biswas,S., Guharoy,M. and Chakrabarti,P. (2008) Structural segments and residue propensities in protein-RNA interfaces: comparison with protein-protein and protein-DNA complexes. *Bioinformation*, **2**, 422–427.
21. Terribilini,M., Sander,J.D., Lee,J.H., Zaback,P., Jernigan,R.L., Honavar,V. and Dobbs,D. (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.*, **35**, W578–W584.
22. Shulman-Peleg,A., Nussinov,R. and Wolfson,H.J. (2009) RsiteDB: a database of protein binding pockets that interact with RNA nucleotide bases. *Nucleic Acids Res.*, **37**, D369–D373.
23. Murakami,Y., Spriggs,R.V., Nakamura,H. and Jones,S. (2010) PiRaNhA: a server for the computational prediction of RNA-binding residues in protein sequences. *Nucleic Acids Res.*, **38**, W412–W416.
24. Kim,O.T., Yura,K. and Go,N. (2006) Amino acid residue doublet propensity in the protein-RNA interface and the application to RNA interface prediction. *Nucleic Acids Res.*, **34**, 6450–6460.
25. Kumar,M., Gromiha,M.M. and Raghava,G.P. (2008) Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins*, **71**, 189–194.
26. Shazman,S. and Mandel-Gutfreund,Y. (2008) Classifying RNA-binding proteins based on electrostatic properties. *PLoS Comp. Biol.*, **4**, e1000146.
27. Shulman-Peleg,A., Shatsky,M., Nussinov,R. and Wolfson,H.J. (2008) Prediction of interacting single-stranded RNA bases by protein binding patterns. *J. Mol. Biol.*, **379**, 299–316.
28. Spriggs,R.V. and Jones,S. (2009) RNA-binding residues in sequence space: Conservation and interaction patterns. *Compu. Biol. Chem.*, **33**, 397–403.
29. Perez-Cano,L. and Fernandez-Recio,J. (2010) Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins*, **78**, 25–35.
30. Ellis,J.J. and Jones,S. (2008) Evaluating conformational changes in protein structures binding RNA. *Proteins*, **70**, 1518–1526.
31. Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
32. Auweter,S.D., Oberstrass,F.C. and Allain,F.H.T. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.*, **34**, 4943–4959.
33. Gupta,A. and Gribskov,M. (2011) The role of RNA sequence and structure in RNA–protein interactions. *J. Mol. Biol.*, **409**, 574–587.
34. Janin,J. (2010) The targets of CAPRI Rounds 13-19. *Proteins*, **78**, 3067–3072.
35. Setny,P. and Zacharias,M. (2011) A coarse-grained force field for protein-RNA docking. *Nucleic Acids Res.*, **39**, 9118–9129.
36. Barik,A., Nithin,C., Manasa,P. and Bahadur,R.P. (2012) A protein-RNA docking benchmark (I): non-redundant cases. *Proteins*, **80**, 1866–1871.
37. Janin,J., Bahadur,R.P. and Chakrabarti,P. (2008) Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.*, **41**, 133–180.

38. Rodier,F., Bahadur,R.P., Chakrabarti,P. and Janin,J. (2005) Hydration of protein-protein interfaces. *Proteins*, **60**, 36–45.
39. Bahadur,R.P., Chakrabarti,P., Rodier,F. and Janin,J. (2004) A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.*, **336**, 943–955.
40. Saha,R.P., Bahadur,R.P. and Chakrabarti,P. (2005) Interresidue contacts in proteins and protein-protein interfaces and their use in characterizing the homodimeric interface. *J. Proteom. Res.*, **4**, 1600–1609.
41. Bahadur,R.P. and Zacharias,M. (2008) The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cell. Mol. Life Sci.*, **65**, 1059–1072.
42. Bernauer,J., Bahadur,R.P., Rodier,F., Janin,J. and Poupon,A. (2008) DiMoVo: a Voronoi tessellation-based method for discriminating crystallographic and biological protein-protein interactions. *Bioinformatics*, **24**, 652–658.
43. Hubbard,S.J. (1992) *NACCESS: Program for Calculating Accessibilities*. Department of Biochemistry and Molecular Biology, University College of London, London, UK.
44. Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
45. McDonald,I. and Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.